



MetaboAnalyst 5.0

A Web-based Tool for Streamlined
Metabolomics Data Analysis

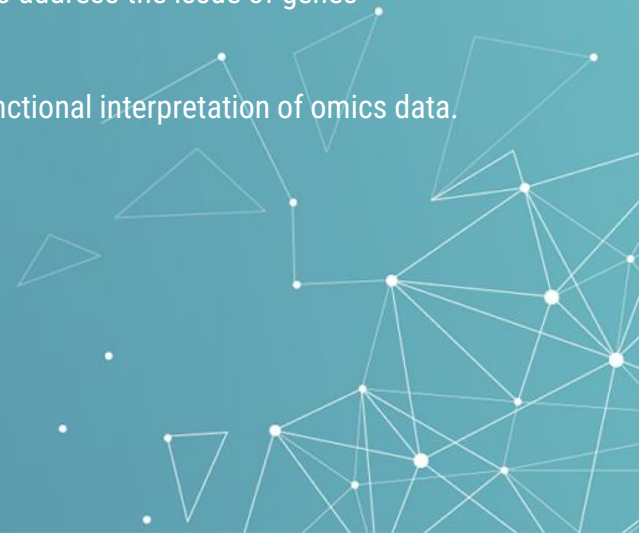
2022.07.12

4. Joint Pathway Analysis

The **Joint Pathway Analysis** module of MetaboAnalyst has added multiple enhancements for Version 5.

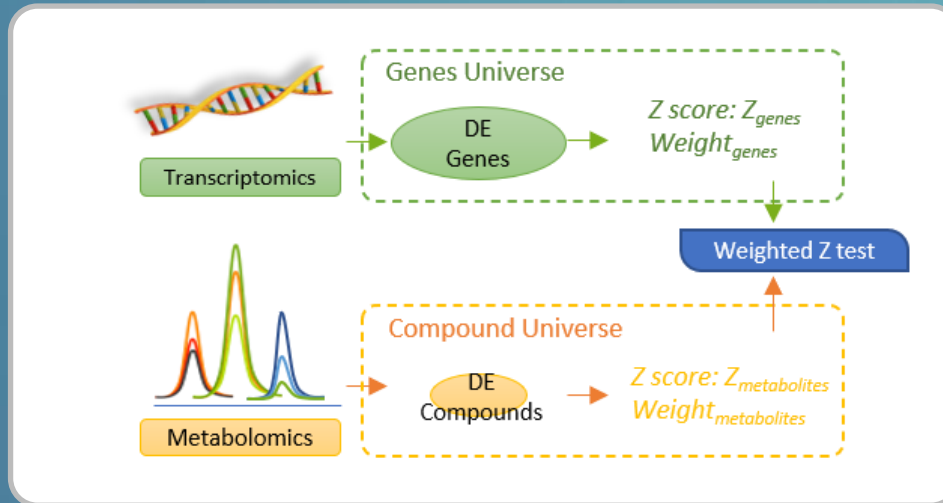
Highlights:

- Added support for an additional 22 model organisms.
- A weighted Z-tests based joint pathway analysis algorithm has been developed to address the issue of genes overwhelming integrated pathway analysis results.
- Multiple integration strategies have been implemented to further enhance the functional interpretation of omics data.
- Support integration for both targeted and untargeted metabolomics.



4.0 Knowledge & Background

- Joint Pathway Analysis is based on weighted integration. The weight strategy for different 'universes' ("transcriptomics data universe" and "metabolomics data universe") are performed with a weighted z-test. The weighted z-test is proposed for combination analysis [by Dmitri V. Zaykin](#) and leveraged here for the weighted integration of different datasets with significantly different sizes. Here we specifically prepared a figure to clearly illustrate the mechanism of this weighted integration of different Omics-data in MetaboAnalyst below.



4.0 Knowledge & Background

Specifically, different weights are assigned based on the proportion of genes and metabolites in the specific 'omics 'universes' to balance the influence from the different sizes of the 'omics inputs upon the integrated pathway results. The adjusted P value is estimated with a weighted Z-test as below,

$$P_{merged} = 1 - \Phi \left(\frac{\sum_{i=1}^2 w_i Z_i}{\sqrt{\sum_{i=1}^2 w_i^2}} \right)$$

Where w_i is the weights of the P values of genes or compounds within individual omics "universe" or "pathway space", respectively; Z_i is the Z score of the corresponding P values of single omics data, usually, $Z_i = \Phi^{-1}(1 - P_i)$; P_i is the P values from the enrichment analysis above; Φ denote the standard normal cumulative distribution function.



4.1 Start Joint Pathway Analysis

MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

Module Overview

| Input Data Type | Available Modules (click on a module to proceed, or scroll down for more details) | | | | | |
|--|---|---------------------|---------------------------------|---------------------------|------------------|-----------------|
| Raw Spectra (mzML, mzXML or mzData) | | | | LC-MS Spectral Processing | | |
| MS Peaks (peak list or intensity table) | | | Functional Analysis | Functional Meta-analysis | | |
| Annotated Features (compound list or table) | | Enrichment Analysis | Pathway Analysis | Joint-Pathway Analysis | Network Analysis | |
| Generic Format (.csv or .txt table files) | Statistical Analysis | Biomarker Analysis | Time-series/Two-factor Analysis | Statistical Meta-analysis | Power Analysis | Other Utilities |

Click here to start

Show R command history

4.2 Integration of genes/proteins with compounds



4.2.1 Data Upload (genes and compounds)

TIP: The Fold change is optional. The titles of the 2 columns need to start with '#'.

Organism:

Metabolomics Type:

Gene list with optional fold changes

| #Official | logFC |
|-----------|-------|
| DEPDC1B | 2.727 |
| CDC6 | 2.309 |
| MELK | 2.65 |
| MCM10 | 2.98 |
| PBK | 3.025 |
| DTL | 2.775 |
| TOP2A | 2.914 |
| RRM2 | 2.535 |
| TYMS | 2.783 |
| CCNA2 | 1.899 |
| CDC25A | 2.571 |
| CDK1 | 2.545 |
| CDC45 | 2.491 |
| BUB1 | 2.732 |
| NUSAP1 | 1.795 |

ID Type:

Compound list with optional fold changes

| #compound | logFC |
|-----------------------|--------|
| Citric Acid | -0.69 |
| Cis-Aconitic Acid | -0.585 |
| Fumaric Acid | -0.433 |
| Stearic acid | -0.602 |
| Tryptophan | -0.828 |
| Malic Acid | -0.331 |
| 2-Ketoisocaproic acid | -0.496 |
| 1,2-Propanediol | -0.352 |
| Salicylic acid | -1.889 |
| 3-Hydroxybutyric acid | -0.717 |
| Trans-Aconitic Acid | -0.412 |
| Inositol | -0.725 |
| Propanoic acid | -0.327 |
| Uric acid | -0.859 |
| Palmitic Acid | -0.535 |

ID Type:

[Try our example](#)

1. Copy and paste your list of genes and metabolites.

2. Specify the ID type.

3. Click "Submit" to upload the data.

4.2.2 Name Matching Results



MetaboAnalyst 5.0 - from raw spectra to patterns and biological insights



Upload

Integrative Analysis

ID list

Set parameter

View result

Download

Exit

The system requires all the IDs (except common compound names) to be matched exactly. The table below shows the matched genes and compounds from the [MetaboAnalyst 5.0](#) compound names, users can further perform [approximate match](#) by clicking the **View** link in the Details column. To **remove** a gene or compound from further analysis, click the **Delete** link in the **Details** column.

Compound Name Mapping

Gene Name Mapping

| Query | Hit | HMDB | KEGG | Details |
|--------|-----------------------------|-----------------------------|------------------------|------------------------|
| C00006 | NADP | HMDB0000217 | C00006 | Delete |
| C00024 | Acetyl-CoA | HMDB0001206 | C00024 | Delete |
| C00026 | Oxoglutaric acid | HMDB0000208 | C00026 | Delete |
| C00029 | Uridine diphosphate glucose | HMDB0000286 | C00029 | Delete |
| C00031 | D-Glucose | HMDB0000122 | C00031 | Delete |
| C00047 | L-Lysine | HMDB0000182 | C00047 | Delete |
| C00049 | L-Aspartic acid | HMDB0000191 | C00049 | Delete |
| C00062 | L-Arginine | HMDB0000517 | C00062 | Delete |
| C00064 | L-Glutamine | HMDB0000641 | C00064 | Delete |
| C00072 | Ascorbate | - | C00072 | Delete |
| C00077 | Ornithine | HMDB0000214 | C00077 | Delete |
| C00084 | Acetaldehyde | HMDB0000990 | C00084 | Delete |
| C00089 | Sucrose | HMDB0000258 | C00089 | Delete |
| C00097 | L-Cysteine | HMDB0000574 | C00097 | Delete |
| C00101 | Tetrahydrofolic acid | HMDB0001846 | C00101 | Delete |
| C00109 | 2-Ketobutyric acid | HMDB0000005 | C00109 | Delete |
| C00111 | Dihydroxyacetone phosphate | HMDB0001473 | C00111 | Delete |
| C00117 | D-Ribose 5-phosphate | HMDB0001548 | C00117 | Delete |
| C00122 | Fumaric acid | HMDB0000134 | C00122 | Delete |
| C00141 | Alpha-ketoglutaric acid | HMDB0000019 | C00141 | Delete |

OK
A total of 389 unique genes were uploaded.

OK
Name matching OK, please inspect (and manual correct) the results then proceed.

Results of the name mapping of the uploaded gene / metabolite data to MetaboAnalyst's internal databases. Scroll down and click "Submit" to continue.

4.2.3 Parameter Selection

1. Select whether to use only metabolic pathways, all pathways, metabolite-only or gene-only pathways.

2. Next, decide on algorithm parameters such as how to combine the gene / metabolite data and how to score important nodes.

3. Scroll down and click "Submit" to perform the Joint Pathway Analysis.

TIP: Please carefully read the instructions of the different options of algorithms before submitting your data.

MetaboAnalyst 5.0 - from raw spectra to patterns and biological insights

Pathway Database

The KEGG pathway was updated in October, 2019 using KEGG API. The integration can occur in two different "universes" defined by **metabolic pathways** or **all pathways**. Metabolic pathways include pathways containing both metabolites and metabolic genes while all pathways include both metabolic pathways as well as gene-only pathways (i.e. regulatory pathways). Users can also perform enrichment analysis for metabolites only using **metabolic pathways (metabolite only)** or for genes only using **all pathways (gene only)**.

Metabolic pathways (integrated)
 All pathways (integrated)
 Metabolic pathways (metabolite only)
 All pathways (gene only)

Algorithm Selection

The topology analysis evaluates the potential importance of a particular molecule (a node) based on its **position** within a pathway. **Degree Centrality** measures the number of links that connect to a node. **Betweenness Centrality** measures the number of shortest paths from all nodes to all the others that pass through a given node. **Closeness Centrality** measures the overall distance from a given node to all other nodes.

For integration methods, there are two general approaches - **tight integration by combining queries** in which genes and metabolites are pooled into a single query and used to perform enrichment analysis within their "pooled universe" or **loose integration by combining p-values** in which enrichment analysis is performed separately for genes and metabolites in their "individual universe", and then individual p-values are combined via **weighted Z-tests**. Moreover, there are three options for computing weights. Let's assume the pathway database contains a total of 100 pathways covering a total of 1000 metabolites and 4000 genes, respectively. Pathway A contains 5 compounds and 45 genes, while pathway B contains 20 compounds and 30 genes.

- **Unweighted** or equal weights (i.e. metabolite: 0.5, gene: 0.5);
- Weights based on the **overall** proportion of each omics within the "universe" (i.e. metabolite: 0.2, gene: 0.8 for all pathways);
- Weights based on the **pathway-level** proportion within individual "pathway space" (i.e. pathway A - metabolite 0.1, gene 0.9; pathway B - metabolite 0.4, gene 0.6)

Note that combining p-values can only be applied to pathways receiving hits from both input types. For pathways with hits from only one input type, p values calculated from their individual universe will be used. In this case, **combining p-values can be viewed as adjusting the confidence level based on new evidence (i.e. input from another omics layer)**. If no new evidence is available, the current confidence level remains.

Enrichment analysis: Hypergeometric Test Fisher's Exact Test

Topology measure: Degree Centrality Betweenness Centrality Closeness Centrality

Integration method: Combine queries
 Combine p values (unweighted)
 Combine p values (overall)
 Combine p values (pathway-level)

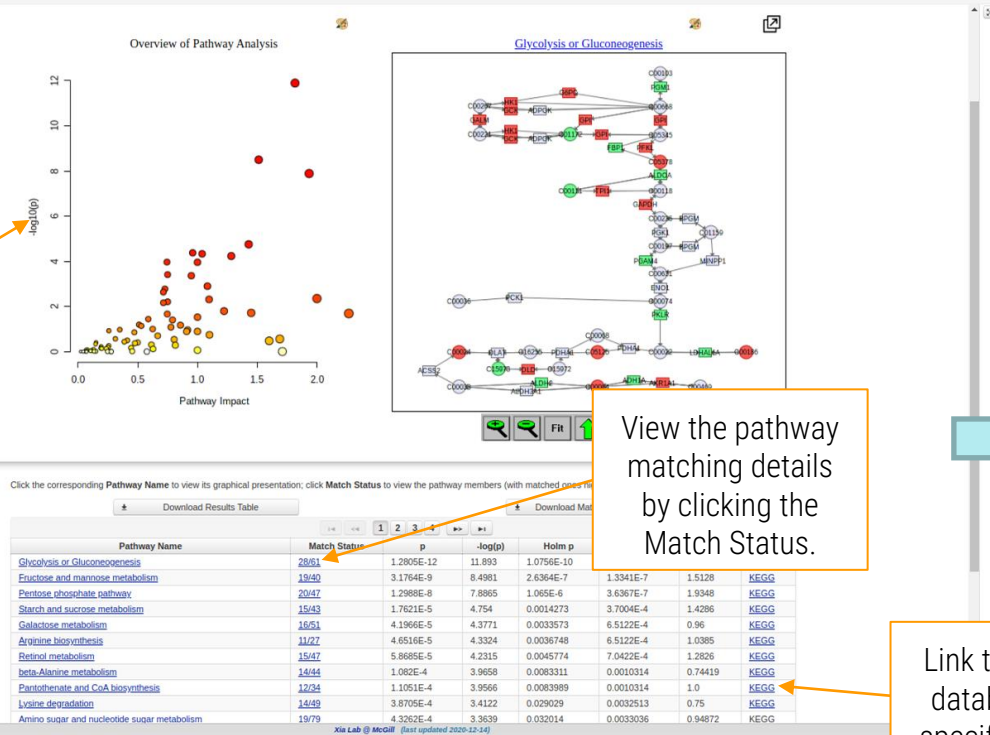
View result
Download
Exit

Submit
Xia Lab @ McGill (last updated 2020-12-14)

4.2.4 Integration Results



MetaboAnalyst 5.0 - from raw spectra to patterns and biological insights



Scatterplot summary of pathway analysis results. Double-click a node to view the corresponding KEGG pathway.

View the pathway matching details by clicking the Match Status.

Matched Features

| Pathway | Members |
|---------------------------------|---|
| Fructose and mannose metabolism | D-Sorbitol; D-Fructose; D-Mannose; beta-D-Fructose 2,6-bisphosphate; D-Mannose 6-phosphate; L-Fucose 1-phosphate; GDP-4-dehydro-6-deoxy-D-mannose; GDP-mannose; D-Mannose 1-phosphate; beta-D-Fructose 6-phosphate; D-Glyceraldehyde 3-phosphate; D-Fructose 1-phosphate; D-Glyceraldehyde; beta-D-Fructose 1,6-bisphosphate; 6-Deoxy-L-galactose; L-Fucosate; alpha-D-Glucose; GDP-L-fucose; Glycerone phosphate; 2-Dehydro-3-deoxy-L-lucosate; AKR1B1, ADR, ALDR1, ALR2, AR, SORD, HEL, S-95n, RDH, SDH, SORD1, XDH, KHK, HK1, HK, HK1-ta, HK1-tb, HK1-tc, HKD, HKI, HMSNR, HXK1, RP79, hexokinase, PFKFB1, F6PK, HL2K, PFRX, MPI, CDG1B, PMI, PM1, PM11, PM11-1, PM1H-22, Sec53, FPGT, GFFP, 1S1A3, FX, P35B, SDR4E1, GMD5, GMD, SDR3E1, GMPPB, LGMDR19, MDDG14, MDDGB14, MDDGC14, PFKL, ATP-PFK, PFK-B, PFK-L, TPI1, HEL, S-49, TIM, TPI, TPID, ALDOA, ALDA, GSD12, HEL, S-87n, TKFC, DAK, NET45, FBP1, FBP, FCSK, 1110046B12RIK, COGF2, FUK, ENOSF1, FUCC, RTS, TYMSAS |

Link to the KEGG database of the specific pathway.

4.3 Integration of genes/proteins with MS peaks



4.3.1 Data Upload (genes and MS peaks)

Organism: Homo sapiens (human)

Metabolomics Type: Untargeted (peak list)

Gene list with optional fold changes

```
#GENE.ID FC
AATK 0.620027928
ABCC5 0.732549041
ABHD3 0.397711093
ABHD4 0.324653754
ACAT1 -0.748452844
ACAT2 -0.655672395
ACLY -0.229198914
ACQ2 -0.449558763
ACSL5 -0.376570988
ADAM19 0.630918962
ADAM8 0.363972481
ADM 0.812138301
ADSL -0.671198436
AHCY -0.415795648
AIMP1 -0.351820758
```

ID Type: Official Gene Symbol

Upload a peak list file

Ion Mode: Positive Mode

Mass Tolerance (ppm): 5.0 (editable)

Retention Time: Not present

Ranked by (1 column only): P values T scores

Mummichog version 2:

Data File: MalariaPeaks.txt 323K

[Try our example](#)

Change the metabolomics type as 'Untargeted (Peak list)'

A peak list must be uploaded here

Click **Submit** to Proceed

4.3.2 Name Matching and peaks processing

The system requires all the IDs (except common compound names) to be matched exactly. The table below shows the matched genes and compounds from the underlying databases. For common compound names, users can further perform [approximate match](#) by clicking the **View** link in the Details column. To **remove** a gene or compound from further analysis, use the **Delete** link in the last column.

Gene Name Mapping

Peaks Processing

Peak Data Integrity Check:

Checking data content ...passed.

A total of 8197 m/z features were found in your uploaded data.

The instrument's mass accuracy is 5 ppm.

The instrument's analytical mode is **positive**.

The uploaded data contains 4 columns.

The column headers of uploaded data are **m.z, p.value, t.score, r.t**.

The range of m/z peaks is trimmed to 50-2000. 0 features have been trimmed.

A total of 8197 input m/z features were retained for further analysis.

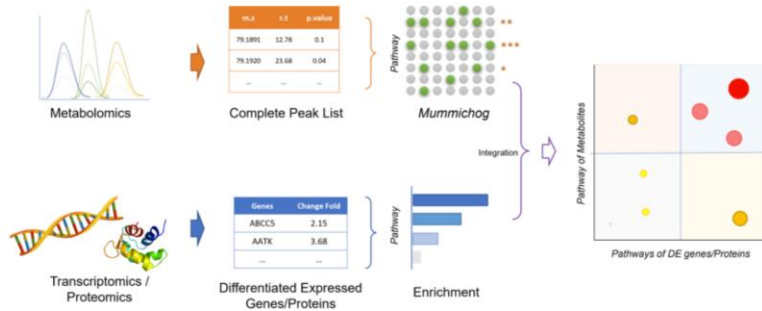
→ Proceed

Results of the name mapping of the uploaded gene to MetaboAnalyst's internal database, and peaks processing results are shown in this page. Click "**Proceed**" to continue.

4.3.3 Parameter Setting

Parameter Settings

For untargeted (global) metabolomics with peaks as input, the [mummichog](#) algorithm ([Li et al](#)) will be used for pathway activity prediction. The resulting pathway p-values will be integrate with pathway p values from [Hypergeometric Test](#) based on significant gene list (see the image below). For advanced users, you can adjust p value cutoff (for selecting significant peaks for mummichog) and/or edit currency metabolites and potential adducts (for peak annotation).



| | |
|---|--|
| Pathway library | <input checked="" type="radio"/> Metabolic Pathways <input type="radio"/> All Pathways |
| View option | <input checked="" type="radio"/> Pathway view <input type="radio"/> Network view |
| Pathway p-value integration | Fisher <input type="button" value="v"/> |
| Advanced options <i>(for mummichog only)</i> | Peak annotation: Currency Compounds Adduct List Sig. peaks cutoff: <input type="text" value="0.04"/> (default select top 10% peaks) |

→ Submit

TIP1: Pathway Library include two options. 'Metabolic Pathways' refers to the pathways include both genes and compounds, while 'All Pathways' also includes the pathways have containing genes only.

TIP2: P value cut-off is estimated automatically as the top 10% significant peaks.

Configure the parameters from this box. Click "**Proceed**" to continue.

4.3.4 Integration Results



The result page is similar to 4.2.4. Different from the compounds uploading, the matched peaks will be shown in dark blue. The result label is showing the putative adduct for this peaks.

Thanks

*If you have any questions please read through the FAQs or contact us at
[Zhiqiang.pang\[at\]xialab.ca](mailto:Zhiqiang.pang@xialab.ca) or [Jeff.xia\[at\]xialab.ca](mailto:Jeff.xia@xialab.ca)*

